



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 16, 2020 – 06:25 PM EST

PDB ID : 6VGE
Title : Crystal structure of the DNA binding domains of human transcription factor ERG, human Runx2 bound to core binding factor beta (Cbfb), in complex with 16mer DNA CAGAGGATGTGGCTTC
Authors : Hou, C.; Tsodikov, O.V.
Deposited on : 2020-01-07
Resolution : 4.25 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.14.6
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.6

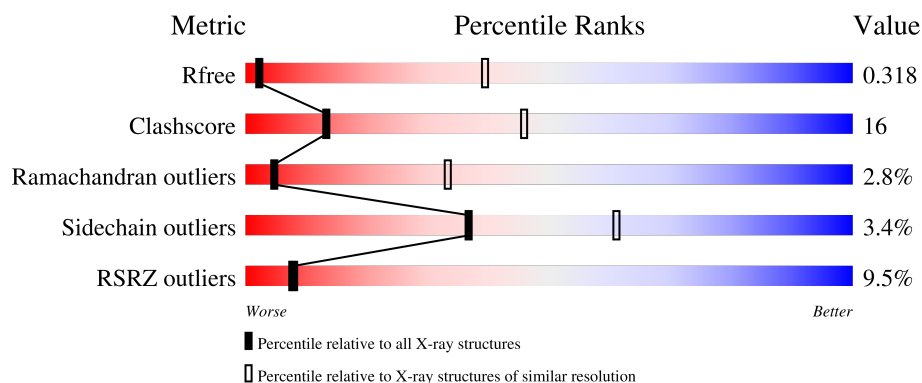
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.25 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1017 (4.72-3.78)
Clashscore	141614	1059 (4.72-3.80)
Ramachandran outliers	138981	1014 (4.72-3.80)
Sidechain outliers	138945	1018 (4.72-3.78)
RSRZ outliers	127900	1072 (4.80-3.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	128	
2	B	16	
3	C	16	
4	D	177	
5	G	156	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3361 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Transcriptional regulator ERG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	98	Total	C	N	O	S	0	0	0
			806	512	143	146	5			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	302	GLY	-	expression tag	UNP P11308
A	303	PRO	-	expression tag	UNP P11308
A	304	HIS	-	expression tag	UNP P11308
A	305	MET	-	expression tag	UNP P11308
A	420	SER	-	expression tag	UNP P11308
A	421	TYR	-	expression tag	UNP P11308
A	422	HIS	-	expression tag	UNP P11308
A	423	ALA	-	expression tag	UNP P11308
A	424	HIS	-	expression tag	UNP P11308
A	425	PRO	-	expression tag	UNP P11308
A	426	GLN	-	expression tag	UNP P11308
A	427	LYS	-	expression tag	UNP P11308
A	428	MET	-	expression tag	UNP P11308
A	429	ASN	-	expression tag	UNP P11308

- Molecule 2 is a DNA chain called DNA (5'-D(P*CP*AP*GP*AP*GP*GP*AP*TP*GP*TP*GP*GP*CP*TP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	16	Total	C	N	O	P	0	0	0
			332	157	62	97	16			

- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*AP*AP*GP*CP*CP*AP*CP*AP*TP*CP*CP*TP*CP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	16	Total	C	N	O	P	0	0	0
			324	154	59	95	16			

- Molecule 4 is a protein called Runt-related transcription factor 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	117	Total	C	N	O	S	0	0	0
			901	567	164	166	4			

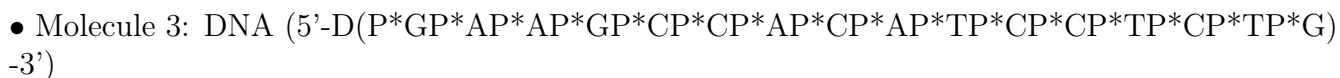
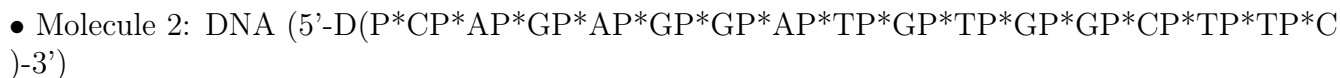
- Molecule 5 is a protein called Core-binding factor subunit beta.

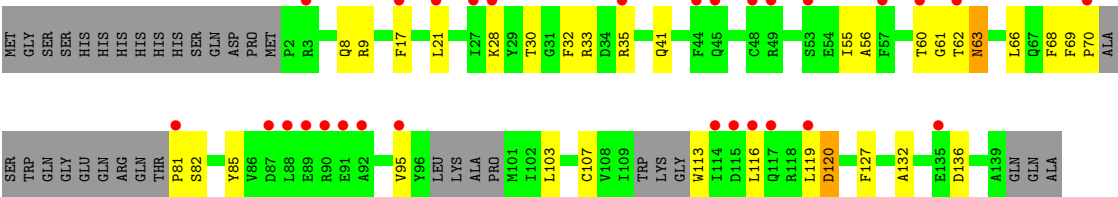
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	G	121	Total	C	N	O	S	0	0	0
			998	622	181	189	6			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-13	MET	-	initiating methionine	UNP Q13951
G	-12	GLY	-	expression tag	UNP Q13951
G	-11	SER	-	expression tag	UNP Q13951
G	-10	SER	-	expression tag	UNP Q13951
G	-9	HIS	-	expression tag	UNP Q13951
G	-8	HIS	-	expression tag	UNP Q13951
G	-7	HIS	-	expression tag	UNP Q13951
G	-6	HIS	-	expression tag	UNP Q13951
G	-5	HIS	-	expression tag	UNP Q13951
G	-4	HIS	-	expression tag	UNP Q13951
G	-3	SER	-	expression tag	UNP Q13951
G	-2	GLN	-	expression tag	UNP Q13951
G	-1	ASP	-	expression tag	UNP Q13951
G	0	PRO	-	expression tag	UNP Q13951

- Molecule 1: Transcriptional regulator ERG





4 Data and refinement statistics

Property	Value	Source
Space group	P 62 2 2	Depositor
Cell constants a, b, c, α , β , γ	103.92Å 103.92Å 322.91Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	35.00 – 4.25 43.69 – 4.24	Depositor EDS
% Data completeness (in resolution range)	95.0 (35.00-4.25) 95.1 (43.69-4.24)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.70 (at 4.28Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.312 , 0.317 0.314 , 0.318	Depositor DCC
R_{free} test set	409 reflections (5.42%)	wwPDB-VP
Wilson B-factor (Å ²)	194.8	Xtriage
Anisotropy	0.477	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 239.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	3361	wwPDB-VP
Average B, all atoms (Å ²)	303.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.82% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.64	0/828	0.77	0/1114
2	B	0.29	0/372	0.79	0/573
3	C	0.30	0/362	0.86	0/555
4	D	0.66	0/920	0.77	0/1253
5	G	0.65	0/1014	0.72	0/1355
All	All	0.59	0/3496	0.77	0/4850

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	806	0	777	36	0
2	B	332	0	181	14	0
3	C	324	0	180	21	0
4	D	901	0	905	19	0
5	G	998	0	956	29	0
All	All	3361	0	2999	100	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 16.

The worst 5 of 100 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:364:LYS:HA	3:C:11:DT:H72	1.40	1.03
1:A:397:GLN:HE22	1:A:402:HIS:HB3	1.43	0.84
1:A:305:MET:HG3	1:A:306:PRO:HD2	1.65	0.79
5:G:28:LYS:HG2	5:G:120:ASP:HB2	1.69	0.74
1:A:329:CYS:HA	1:A:342:THR:HB	1.69	0.74

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	96/128 (75%)	79 (82%)	12 (12%)	5 (5%)	2	22
4	D	115/177 (65%)	93 (81%)	18 (16%)	4 (4%)	3	29
5	G	113/156 (72%)	104 (92%)	9 (8%)	0	100	100
All	All	324/461 (70%)	276 (85%)	39 (12%)	9 (3%)	5	33

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	342	THR
1	A	386	TYR
4	D	163	ASN
4	D	135	THR
4	D	150	ASP

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was

analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	85/113 (75%)	82 (96%)	3 (4%)	36	60
4	D	100/156 (64%)	96 (96%)	4 (4%)	31	57
5	G	106/135 (78%)	103 (97%)	3 (3%)	43	65
All	All	291/404 (72%)	281 (97%)	10 (3%)	37	61

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	141	LYS
4	D	190	ARG
5	G	63	ASN
4	D	134	LYS
4	D	222	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
5	G	8	GLN
5	G	134	GLN
5	G	63	ASN
1	A	376	ASN
5	G	67	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	98/128 (76%)	-0.10	2 (2%) 65 56	213, 252, 281, 296	0
2	B	16/16 (100%)	-0.17	0 100 100	251, 272, 299, 304	0
3	C	16/16 (100%)	-0.32	0 100 100	239, 278, 331, 335	0
4	D	117/177 (66%)	0.06	4 (3%) 45 36	234, 274, 305, 322	0
5	G	121/156 (77%)	1.17	29 (23%) 0 1	354, 382, 421, 432	0
All	All	368/493 (74%)	0.35	35 (9%) 8 8	213, 280, 412, 432	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	G	53	SER	5.3
5	G	88	LEU	5.3
5	G	45	GLN	4.8
5	G	44	PHE	4.3
5	G	48	CYS	4.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.